

AN: / PAT 1988-300842
TI: * Checking correct and consistent use of units or chemical
formulae checking whether strings represent allowable notation
✓ of allowable unit in text processing system
PN: **EP287713-A**
PD: 26.10.1988
AB: A list of symbols or set of prodn. rules that represent
units is stored. Criteria, e.g. typical features of the
notation or of the context of units, that state whether or not
a symbol or a combination of symbols is interpreted as a unit,
are stored. Allowable units and notations are specified. The
text is searched for strings that include any of the listed
symbols and can be interpreted as a unit or part of a unit
pursuant to the stored criteria. If such a string is found, the
related unit and the complete notation are compared. An error
signal is generated if the unit or the notation is not
allowable.; Automatically detecting at least most frequently
made mistakes.
PA: (CHEZ) OCE NEDERLAND BV;
IN: CORNELISSEN P A M; VAN GASTEREN T H; GASTEREN V; VAN
GASTEREN T H I E;
FA: **EP287713-A** 26.10.1988; DE3750135-G 28.07.1994;
EP287713-B1 22.06.1994; US5159552-A 27.10.1992;
CO: AT; BE; CH; DE; EP; ES; FR; GB; GR; IT; LI; LU; NL; SE; US;
DR: AT; BE; CH; DE; ES; FR; GB; GR; IT; LI; LU; NL; SE;
IC: G06F-015/20; G06F-015/38;
MC: T01-J09;
DC: T01;
FN: 1988300842.gif
PR: EP0200768 23.04.1987;
FP: 26.10.1988
UP: 28.07.1994



03P19154



⑬ BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENTAMT

⑫ Übersetzung der
europäischen Patentschrift

⑤① Int. Cl.⁵:
G 06 F 15/20

⑧⑦ EP 0 287 713 B1

⑩ DE 37 50 135 T 2

②① Deutsches Aktenzeichen: 37 50 135.6
⑧⑥ Europäisches Aktenzeichen: 87 200 768.7
⑧⑥ Europäischer Anmeldetag: 23. 4. 87
⑧⑦ Erstveröffentlichung durch das EPA: 26. 10. 88
⑧⑦ Veröffentlichungstag
der Patenterteilung beim EPA: 22. 6. 94
④⑦ Veröffentlichungstag im Patentblatt: 8. 12. 94

⑦③ Patentinhaber:
Oce-Nederland B.V., Venlo, NL

⑦④ Vertreter:
ter Meer, N., Dipl.-Chem. Dr.rer.nat.; Müller, F.,
Dipl.-Ing., 81679 München; Steinmeister, H.,
Dipl.-Ing.; Wiebusch, M., 33617 Bielefeld; Urner, P.,
Dipl.-Phys. Ing.(grad.); Merkle, G., Dipl.-Ing. (FH),
Pat.-Anwälte, 81679 München

⑧④ Benannte Vertragsstaaten:
DE, FR, GB, IT, NL, SE

⑦② Erfinder:
van Gasteren, Theodrikus Henricus I.L., NL-5976 PR
Kronenberg, NL; Cornelissen, Petrus Antonius
Maria, NL-5911 AC Venlo, NL

⑤④ Textverarbeitungssystem und -verfahren zur Prüfung des richtigen und folgerichtigen Gebrauchs von Einheiten und chemischen Formeln in einem Textverarbeitungssystem.

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patentamt inhaltlich nicht geprüft.

DE 37 50 135 T 2

DE 37 50 135 T 2

P 37 50 135.6-08

Die vorliegende Erfindung betrifft das Gebiet der automatischen Textverarbeitung. Insbesondere befaßt sich die Erfindung mit einem Textverarbeitungssystem und Verfahren zur automatischen Feststellung einer bestimmten
5 Art von Fehlern, die in einem Verstoß gegen bestimmte Regeln über den Gebrauch von Einheiten bestehen, in einem gegebenen Text.

In einigen herkömmlichen Textverarbeitungssystemen ist es möglich,
10 Schreibfehler mit Hilfe eines sogenannten Rechtschreibprozessors festzustellen. Der Rechtschreibprozessor überprüft jedes Wort des Textes anhand eines Wörterverzeichnisses. Wenn ein Wort des Textes nicht mit irgendeinem Wort des Wörterverzeichnisses identifiziert werden kann, so wird geschlossen, daß dieses Wort falsch geschrieben ist. Einige Rechtschreibprozessoren
15 sind sogar in der Lage, den Schreibfehler automatisch zu korrigieren, indem das falsch geschriebene Wort durch ein Wort aus dem Wörterverzeichnis ersetzt wird. Um das richtige Wort im Wörterverzeichnis herauszufinden, sind ausgefeilte Berechnungsmethoden entwickelt worden, die auf einer Analyse von Ähnlichkeiten zwischen den Wörtern und/oder einem statistischen Ansatz unter Berücksichtigung der verschiedenen Auftretishäufigkeiten bestimmter Arten von Schreibfehlern oder gar auf einer Analyse des Kontextes
20 des falsch geschriebenen Wortes unter Ausnutzung der Redundanz des Textes basieren.

25 Aus Research Scientific and Application Industry, Band 10, Nr. 3, Oktober 1957, Seiten 383-389, I.K. Belskaja: "Machine Translation of Languages" ist ein Satzanalysesystem bekannt, das verschiedene Arten von Wörterverzeichnissen aufweist, die entsprechend dem Quellentext herangezogen werden. Darin hat ein bestimmtes zu übersetzendes Wort nur eine Bedeutung und
30 wird unzweideutig als ein zu übersetzendes Wort interpretiert.

Aus EP-0 093 249 ist ein System zur Feststellung und Korrektur von kontextbezogenen Fehlern in einem Textverarbeitungssystem bekannt, bei dem Sätze von fehlerträchtigen Datensegmenten zusammen mit Kontextmerkmalen
35 gespeichert sind. Der jedes fehlerträchtige Datensegment umgebende Text in einem Dokument wird mit den definierten Kontextmerkmalen verglichen, und im Fall der Nichtvereinbarkeit wird das fehlerträchtige Datensegment

hervorgehoben. Auf der Anzeige wird ein Satz von auf das hervorgehobene Datensegment bezogenen Datensegmenten angezeigt.

- 5 In Verbindung mit der Verwendung von Einheiten und Größen, beispielsweise in einem wissenschaftlichen Text, wird jedoch häufig eine Anzahl spezieller Probleme und entsprechend eine Anzahl spezieller Fehler und Unrichtigkeiten angetroffen, die von den bekannten Rechtschreibprozessoren nicht behandelt werden können.
- 10 Vor einer detaillierten Analyse dieser Arten von Fehlern, die in Verbindung mit Einheiten auftreten können, erscheint es zweckmäßig, klare Definitionen für die Begriffe der Einheit und Größe anzugeben, wie sie in dieser Beschreibung verstanden werden.
- 15 Eine Größe ist definiert als ein Merkmal oder eine Eigenschaft eines Objektes, die quantitativ beschrieben werden kann. Solch eine quantitative Beschreibung ist das Ergebnis der Bestimmung des Verhältnisses zwischen der Größe und einer bestimmten Bezugsgröße. Die Bezugsgröße wird "Einheit" genannt, und das Verhältnis wird "Zahlenwert" der Größe genannt.
- 20 Größen können nach Arten von Größen gruppiert werden. Zwei Größen sind definitionsgemäß von derselben Art, wenn sie miteinander verglichen werden können, d.h., wenn sie mit derselben Einheit verglichen werden können. Das Produkt oder der Quotient von zwei Größen derselben Art oder verschiedener Arten ergibt eine neue Größe einer anderen Art. Somit gibt es eine
- 25 Vielzahl von mathematischen Beziehungen zwischen verschiedenen Arten von Größen und zwischen den zugehörigen Einheiten. Innerhalb eines speziellen Anwendungsgebietes ist es möglich, einen Satz von Basisgrößen in solcher Weise zu bestimmen, daß alle in diesem Anwendungsgebiet auftretenden Größen
- 30 als ein Produkt geschrieben werden können, das aus Basisgrößen mit bestimmten positiven oder negativen Potenzen besteht. Dieses Produkt (oder der Satz der Potenzen jeder Basisgröße in diesem Produkt) wird Dimension der Größe genannt.
- 35 Ein Einheitensystem wird aufgestellt, indem Basiseinheiten für die Basisgrößen bestimmt werden, und unter der Dimension einer Einheit wird die Dimension der zugehörigen Art von Größen verstanden.

Wenn als Beispiel zur Illustration der oben gegebenen Definitionen die Größe der Kraft

5 N (5 Newton)

5 betrachtet wird, so ist die Art der Größe "Kraft", der Zahlenwert ist 5, die Einheit ist Newton und die Dimension der Einheit ist $\text{kg} \cdot \text{m} \cdot \text{s}^{-2}$, wobei kg, m und s die Basis-SI-Einheiten für Masse, Länge und Zeit sind.

10 Obgleich das obige Beispiel aus dem Gebiet der Naturwissenschaften entnommen wurde, versteht es sich, daß der Begriff der Einheiten und die Anwendbarkeit der vorliegenden Erfindung nicht auf das Gebiet der Naturwissenschaften beschränkt ist. Die Grundgedanken der vorliegenden Erfindung können in jedem Anwendungsgebiet angewandt werden, in dem quantitative Beschreibungen verwendet werden, beispielsweise im Finanzwesen (Währungseinheiten), in Kochbüchern, bei der Lagerverwaltung (Anzahl von Artikeln pro Schachtel) usw..

20 In einem gegebenen Anwendungsgebiet existiert im allgemeinen eine Vielzahl von verschiedenen Einheitensystemen, und jede Art von Größe kann durch eine große Vielzahl unterschiedlicher Einheiten beschrieben werden. Die Verwendung von Einheiten und Einheitensystemen wird durch verschiedene Normen geregelt. Während einige dieser Normen nur in bestimmten geographischen Regionen oder bestimmten Anwendungsgebieten anwendbar sind, beruhen andere Normen wie etwa die SI-Norm auf einem internationalen Übereinkommen und haben einen breiten Anwendungsbereich. Die Verwendung der SI-Norm ist von vielen nationalen und internationalen Behörden vorgeschrieben. Nach dieser Norm sind einige ältere Einheiten verboten, und einige sind nur in bestimmten wissenschaftlichen Disziplinen erlaubt. Jedoch werden viele Einheiten oder Einheitensysteme, die nach der SI-Norm verboten sind, nichtsdestoweniger in der Praxis häufig benutzt.

30 Wenn für einen Text die Verwendung einer bestimmten Norm vorgeschrieben ist, können folglich viele Fehler auftreten, die in der Verwendung einer verbotenen Einheit bestehen, während die Schreibweise des Namens oder des Symbols für die Einheit korrekt sein mag. Wenn eine Größe, die mit Hilfe einer verbotenen Einheit beschrieben wurde, in einer zulässigen Einheit ausgedrückt werden soll, ist es im allgemeinen notwendig, den Zahlenwert (oder die Zahlenwerte) gemäß einer bestimmten Umwandungsformel zu

ändern.

- Eine weitere wichtige Art von Fehlern betrifft die Bezeichnung der Einheiten, die ebenfalls der gegebenen Norm entsprechen muß. Im Gegensatz zu einem normalen Wort kann eine einzelne Einheit auf zahlreiche verschiedene Weisen bezeichnet werden. Die Basiseinheiten und einige abgeleitete Einheiten können entweder durch einen speziellen Namen oder ein spezielles Symbol bezeichnet werden. Die abgeleiteten Einheiten können außerdem durch verschiedene Kombinationen von Namen oder Symbolen bezeichnet werden.
- 5 Im Fall einer Kombination von Symbolen kann auch das Layout variiert werden.

Beispielsweise geben die folgenden Ausdrücke alle dieselbe Einheit, nämlich die Einheit "Watt" an.

15

- (a) W
- (b) Js^{-1}
- (c) J s^{-1}

(Es ist zu beachten, daß (b) und (c) in einem Textverarbeitungssystem nicht automatisch als identisch erkannt werden!)

20

- (d) J/s
- (e) N·m/s
- (f) $\text{m}^2\text{kg s}^{-3}$
- (g) Watt
- 25 (h) Joule pro Sekunde

Im allgemeinen sind nur einige der möglichen Bezeichnungen nach der gegebenen Norm zulässig oder gebräuchlich. Außerdem sollten die Regeln zur Bestimmung des Layouts für den gesamten Text konsistent sein. Manchmal müssen sogar Mehrdeutigkeiten in der Bezeichnung aufgelöst werden (z.B.: kg/(ha·yr) anstelle von kg/ha·yr).

30

Eine weitere spezielle Art von Fehlern betrifft die Verwendung von Präfixen wie etwa "m" für "milli", "k" für "kilo" und dergleichen. Beispielsweise muß das Auftreten von mehr als einem Präfix in Verbindung mit demselben Symbol vermieden werden, und die Verwendung von Präfixen sollte an die Größenordnung des Zahlenwertes angepaßt sein, um unhandliche Ausdrücke

35

wie 0,0001 km zu vermeiden.

In einigen Fällen ist die Dimension einer in dem Text verwendeten Einheit nicht mit der zu beschreibenden Größe vereinbar, oder die Dimension existiert in dem Anwendungsgebiet nicht.

Ein weiteres Beispiel für Fehler, die in Verbindung mit Einheiten auftreten, ist die Vermischung von Symbolen zur Bezeichnung der Einheit mit dem Namen des zu messenden Objekts (z.B.: "kg Trockenmasse pro ha"). Um diese Fehler zu beseitigen, muß im allgemeinen der gesamte Satz geändert werden.

Wenn Fehler der oben als Beispiel angegebenen Arten in einem Dokument festgestellt und korrigiert werden sollen, so muß der Text vom Herausgeber oder Autor durchgesehen werden. Dies ist eine zeitraubende und mühsame Beschäftigung.

Es ist deshalb eine erste Aufgabe der Erfindung, ein Textverarbeitungssystem und Verfahren zur automatischen Feststellung wenigstens der häufigsten dieser Fehler zu schaffen.

Gemäß der Erfindung wird diese Aufgabe durch das in Anspruch 1 angegebene Verfahren gelöst.

Die in Anspruch 1 angegebenen Schritte a) bis c) sind vorbereitende Schritte, die ausgeführt werden müssen, bevor ein Text oder eine Anzahl von Texten in dem Textverarbeitungssystem zu überprüfen ist. Die Information, die zum Auffinden des Ortes einer Angabe einer Einheit in dem Text, zur Interpretation der Angabe und zum Identifizieren der Einheit benötigt werden, wird in Schritten a) und b) gespeichert.

In Schritt a) wird eine Liste von Symbolen gespeichert, die zur Bezeichnung einer Einheit verwendet werden können. Wenn nur die Bezeichnung von Einheiten zu überprüfen ist, reicht es aus, die Symbole der zulässigen Einheiten zu speichern. Wenn verbotene Einheiten in dem Text zu beseitigen sind, müssen auch die Symbole für verbotene Einheiten gespeichert werden. Wahlweise können zusätzlich zu den Symbolen die vollständigen Namen der Einheiten gespeichert werden.

- Im allgemeinen ist zur zuverlässigen Interpretation der Angaben von Einheiten zusätzliche Information erforderlich. Beispielsweise kann es vom Kontext abhängen, ob ein Symbol tatsächlich für eine Einheit steht oder ob es ein gewöhnliches Wort oder Teil eines gewöhnlichen Wortes ist. Da außerdem eine
- 5 Einheit häufig durch einen Ausdruck bezeichnet werden wird, der aus mehreren Symbolen oder Namen von Einheiten besteht, ist es notwendig, alle Symbole oder Namen, die zu diesem Ausdruck gehören, zu erfassen und die Beziehung zwischen ihnen zu analysieren. Die Kriterien für die Entscheidung, ob ein Symbol eine Einheit bezeichnet - allein oder in Kombination mit anderen
- 10 Symbolen - und für das Identifizieren der Einheit werden in Schritt b) gespeichert. Die Auswahl dieser Kriterien ist von dem Anwendungsgebiet abhängig. Es ist jedoch möglich, einige typische Beispiele für nützliche Kriterien anzugeben:
- 15 - Es ist möglich, Namen oder Symbole zu spezifizieren, die stets als Einheiten interpretiert werden müssen, wenn sie allein stehen oder mit einem Präfix kombiniert sind.
- 20 - Einheiten in Formeln werden häufig in Drucktypen angegeben, die von den für den Text verwendeten Drucktypen verschieden sind.
- Wenn eine Zeichenkette, d.h., eine Folge von Schriftzeichen zwischen Leerzeichen, nur aus Symbolen für Einheiten besteht und kein gewöhnliches Wort ist, das in einem Wörterverzeichnis gefunden werden kann, so
- 25 kann sie als das Produkt der durch die einzelnen Symbole angegebenen Einheiten interpretiert werden.
- Wenn einzelne Symbole (die von einem Exponenten begleitet sein können) durch Leerzeichen, Malpunkte ("," oder "x"), einen Schrägstrich ("/") oder dergleichen getrennt sind, oder wenn Namen von Einheiten
- 30 durch das Wort "mal" oder "pro" getrennt sind, so kann in ähnlicher Weise der gesamte Ausdruck als ein Produkt oder ein Quotient der einzelnen Einheiten interpretiert werden.
- 35 - Weitere leistungsfähige Interpretationskriterien können sich aus der syntaktischen Struktur des Textes ergeben. Wie im Stand der Technik bekannt ist, kann die syntaktische Struktur eines Satzes mit Hilfe eines

- sogenannten Parsers analysiert werden, der die Bestandteile eines Satzes und die syntaktische Kategorie der einzelnen Bestandteile bestimmt. Wegen der Beschränkungen für mögliche syntaktische Kategorien für eine Einheit kann für viele Zeichenketten des Satzes entschieden werden, daß sie nicht als eine Einheit interpretiert werden dürfen. Andererseits kann ein Symbol oder eine Kombination von Symbolen positiv mit einer Einheit identifiziert werden, wenn sie syntaktisch direkt auf eine Zahlenangabe (die den Zahlenwert der Größe angibt) oder auf ein die Größe angegebendes Stichwort bezogen ist. In den meisten Fällen genügen die Positionen von Angaben von Größen, Zahlen und Einheiten in einem Satz einem verhältnismäßig einfachen Schema (z.B.: "Eine Masse von 1 kg unterliegt einer Gravitationskraft zwischen 9.79 und 9.82 N"). Es ist deshalb möglich, einfache Kriterien aufzustellen (z.B., daß eine Zahl dem Symbol für die Einheit unmittelbar vorausgeht), die überprüft werden können, ohne daß viel Rechenzeit verbraucht wird. Wenn andererseits die vollständige syntaktische Struktur eines Satzes mit Hilfe eines Parsers analysiert wird, ist es sogar möglich, verhältnismäßig komplexe Strukturen korrekt zu interpretieren. Zum Beispiel in dem Satz:
- "Spitzen-Ausbringungsraten reichen von bis zu 0,012 im Herbst bis nur 0,002 kg N/ha/d im Winter".
- Es ist zu erkennen, daß die Größe die Ausbringungsrate (von Stickstoff (N)) ist und daß die zu den beiden Zahlen 0,012 und 0,002 gehörende Einheit kg/(ha·d) ist.
- In Schritt c) wird spezifiziert, welche Einheiten und Bezeichnungen von Einheiten zulässig sind. Vorzugswise werden diese Spezifikationen in Übereinstimmung mit einer oder mehreren offiziellen oder inoffiziellen Normen gemacht, und die entsprechende Information wird so gespeichert, daß der Benutzer im Normalbetrieb nur zwischen verschiedenen Normen wählen kann, aber nicht die normgemäßen Vorschriften ändern kann.

Um die zulässigen Einheiten zu spezifizieren, ist es möglich, eine Liste von Symbolen der zulässigen Einheiten zu speichern oder für jedes der in Schritt a) gespeicherten Symbole einen logischen Wert zu definieren. Alternativ können alle Einheiten sowie zugehörige Informationen (Zulässigkeit, mögliche Bezeichnungen, Dimension, Art der Größe und dergleichen) in dem Textverarbeitungssystem in einer internen Repräsentation gespeichert werden.

Außerdem kann ein Satz von Regeln für den Aufbau zulässiger Bezeichnungen für abgeleitete Einheiten gespeichert werden. Zusätzliche Information, die in Schritt c) spezifiziert werden kann, bezieht sich auf die Verwendung von Präfixen und das gewünschte Layout der Bezeichnung von Einheiten. Diese Information kann in solcher Weise gespeichert werden, daß der Benutzer frei zwischen verschiedenen Layouttypen wählen kann, die nach der ausgewählten Norm zulässig sind.

Nachdem die oben beschriebenen vorbereitenden Schritte abgeschlossen sind, wird ein gegebener Text überprüft, indem die in Anspruch 1 angegebenen Schritte d) bis f) aufgeführt werden.

In Schritt d) wird der Text unter Verwendung der in Schritten a) und b) gespeicherten Informationen nach Angaben von Einheiten durchsucht.

In einem direkten Zugang kann dies dadurch geschehen, daß aufeinanderfolgend jede Zeichenkette des Textes untersucht wird. Dies hat den Vorteil, daß jede Angabe einer Einheit mit Sicherheit festgestellt werden kann.

Andererseits wird es in vielen Anwendungen vorzuziehen sein, einen indirekten Zugang zu wählen.

Gemäß einer bevorzugten Ausführungsform der Erfindung wird der Text zunächst nach Zahlen (Zeichenketten, die nur aus Ziffern bestehen) durchsucht, und nur die Zeichenketten unmittelbar hinter einer Zahl werden daraufhin überprüft, ob sie als eine Einheit interpretiert werden können. Eine statistische Analyse hat ergeben, daß durch dieses Verfahren in einem durchschnittlichen Text mehr als 95 % aller Angaben von Einheiten erfaßt werden können. Da Zahlen in einem Textverarbeitungssystem recht einfach festgestellt werden können, führt dieses indirekte Suchverfahren zu einer drastischen Verringerung der Verarbeitungszeit. In speziellen Anwendungsgebieten, beispielsweise im Finanzwesen, kann es auch erforderlich sein, die Zeichenketten unmittelbar vor der Zahl zu überprüfen.

Alternativ kann der Text zunächst nach Angaben von Größen durchsucht werden, indem die Zeichenketten mit einer Tabelle von Schlüsselwörtern und Symbolen oder Erzeugungsregeln, die sich auf Größen beziehen, verglichen

werden. Dann brauchen nur die Zeichenketten in einer bestimmten Umgebung der Größe darauf überprüft zu werden, ob sie Einheiten bezeichnen.

- 5 Gemäß einem weiteren indirekten Zugang werden die Zeichenketten des Textes mit einem Wörterverzeichnis verglichen, und wenn eine Zeichenkette als ein "normales" Wort identifiziert wird, wird nicht versucht, die Zeichenkette als eine Einheit zu interpretieren. Dieser Ansatz ist besonders vorteilhaft, wenn das erfindungsgemäße Verfahren mit der Verwendung eines herkömmlichen Rechtschreibprozessors kombiniert wird. In diesem Fall kann
- 10 das Wörterverzeichnis des Rechtschreibprozessors verwendet werden, sofern spezielle Vorkehrungen für diejenigen Wörter des Wörterverzeichnisses getroffen werden, die dieselbe Schreibweise wie ein Name oder ein Symbol einer Einheit oder einer Kombination/Zusammensetzung von Einheiten haben.
- 15 Wenn eine Angabe einer Einheit in Schritt d) lokalisiert und interpretiert worden ist, wird in Schritt e) überprüft, ob die Einheit und/oder die Bezeichnung derselben zulässig ist. Dies wird allgemein bewerkstelligt, indem überprüft wird, ob die in Schritt c) spezifizierten Regeln erfüllt sind. Zum Beispiel wird zunächst überprüft, ob die Einheit, die durch die in Schritt d) gefundene
- 20 ne Zeichenkette bezeichnet wird, eine Dimension hat, die tatsächlich in dem Anwendungsgebiet (z.B. der Natur) auftritt und/oder mit der Dimension der Größe vereinbar ist, auf die sie in dem Text bezogen ist. Dann wird eine geeignete zulässige Einheit derselben Dimension ausgewählt und eine zulässige Bezeichnung derselben aufgebaut. Schließlich wird überprüft, ob die aufgebaute
- 25 te Bezeichnung mit der im Text gefundenen identisch ist.

Falls die Überprüfung in Schritt e) einen Fehler ergibt, wird in Schritt f) ein Fehlersignal erzeugt.

- 30 Bei einer speziellen Ausführungsform des Verfahrens nach Anspruch 1 können in Schritt a) nur die verbotenen Einheiten gespeichert werden. In diesem Fall wird in Schritt c) festgelegt, daß alle Einheiten verboten sind. Der Schritt e) ist dann trivial, und ein Fehlersignal wird erzeugt, sobald in Schritt d) eine Angabe einer verbotenen Einheit gefunden wird.

35

Das Fehlersignal kann dem Benutzer angezeigt werden, so daß von dem Benutzer eine notwendige Korrektur vorgenommen werden kann.

Vorzugsweise wird jedoch eine vorgeschlagene Korrektur ebenfalls dem Benutzer angezeigt und/oder automatisch ausgeführt. Wenn die Einheit durch eine andere ersetzt wird, wird auch überprüft, welche Zahl oder Zahlen zu dieser Einheit gehören, und der Zahlenwert dieser Größe wird in Übereinstimmung mit einer vorgegebenen Umwandlungsformel angepaßt. Wenn eine Syntaxanalyse ausgeführt worden ist, kann die darin enthaltene Information ausgewertet werden, um die Zahlen zu identifizieren, die zu der Einheit gehören.

- 5
- 10 Ein bevorzugtes Ausführungsbeispiel der Erfindung wird nachfolgend im Zusammenhang mit der Zeichnung erläutert.

Die einzige Zeichnung ist ein Flußdiagramm, das ein Verfahren zur Überprüfung, ob die Verwendung von Einheiten und Größen in einem Text der SI-Norm entspricht, illustriert.

In dem gezeigten Ausführungsbeispiel wird angenommen, daß Hintergrundinformation bezüglich der Symbole für Einheiten, zulässiger Einheiten, zulässiger Dimensionen, Interpretationskriterien, Formeln zur Umwandlung von Einheiten, Schlüsselwörtern, Symbolen und Dimensionen von Größen und dergleichen vorab gespeichert worden ist.

1. Optionen Bemerkung: dies ist ein Beispiel (SI)

- 25 In Schritt 10 wird eine Anzahl von Optionen durch den Benutzer ausgewählt. die verfügbaren Optionen beziehen sich auf die folgenden Probleme:

1.1 Die Sprache des Dokuments.

- 30 Diese Information wird benötigt zur Interpretation von Schlüsselwörtern und zur Handhabung von Besonderheiten der Sprache wie etwa die Notation von Zahlen. Weiterhin kann die Liste von Einheiten verkürzt werden, weil einige verbotene Einheiten in bestimmten Sprachen nicht auftreten.

- 35 1.2 Vollständige Namen von Einheiten.

In dem meisten Fällen muß eine vollständig ausgeschriebene Einheit durch

ihr Symbol ersetzt werden. Manchmal ist jedoch ein Name zulässig (z.B. 5 Jahre).

1.3 Das Layout der der Einheiten.

5

Es gibt drei Arten der Bezeichnung von Einheiten. Eine von ihnen muß konsistent in dem gesamten Dokument verwendet werden.

1. Ein Punkt als Trenner und ein Schrägstrich ("/") im Fall der Division.
Beispiel: kg/(ha·yr)
2. Ein Punkt als Trenner und negative Potenzen im Fall einer Division.
Beispiel: kg·ha⁻¹·yr⁻¹.
3. Ein Leerzeichen als Trenner und negative Potenzen im Fall einer Division.
Beispiel: kg ha⁻¹ yr⁻¹.

1.4 Das Layout der Potenzen.

20

1.5 Die Bezeichnung von Prozentwerten.

Das %-Zeichen kann an die Zahl angehängt oder von dieser durch ein Leerzeichen getrennt sein. Einer dieser Modi muß in dem Dokument konsistent angewandt werden. Wenn ein Fehler auftritt, wird dieser ohne jegliche Erwähnung korrigiert.

2. Suche nach Angaben von Einheiten

Um den Ort einer Einheit in dem Text zu finden, wird der Text zunächst nach Zahlen durchsucht (Schritte 12-16) und anschließend wird überprüft, ob die nächste Zeichenkette oder die nächsten Zeichenketten als Einheiten interpretiert werden können (Schritte 18 bis 26).

2.1 Suche nach Zahlen

In Schritt 12 werden die einzelnen Zeichenketten des Textes eine nach der

anderen durchgegangen, und es wird überprüft, ob eine Zeichenkette eine Zahl darstellt. Es wird angenommen, daß eine Zeichenkette eine Zahl darstellt, wenn sie Ziffern, aber keine Buchstaben enthält (z.B.: 1.735,8). Einige wenige Ausnahmen (spezielle Buchstabensequenzen zur Darstellung spezieller
5 Zeichen gemäß dem SGML-Standard) sind zugelassen.

In einer abgewandelten Ausführungsform ist der Begriff der Zahl verallgemeinert, so daß Wörter oder Buchstabensequenzen, die sich auf Zahlen oder Mengen beziehen, wie etwa "fünf", "Million", "ein Dutzend", "einige" ...
10 ebenfalls erkannt werden.

Um den Wert der Zahl festzustellen (für etwaige Umwandlungen) muß die Möglichkeit berücksichtigt werden, daß eine Zahl mehr als eine Zeichenkette umfaßt (z.B. 60.000). Aus diesem Grund wird überprüft, ob die nächste
15 Zeichenkette mit einer Ziffer beginnt.

Ein spezielles Problem im Zusammenhang mit der Bestimmung des Wertes einer Zahl besteht in der Mehrdeutigkeit in der Notation von Tausendern und Dezimalstellen. In der englischen Sprache wird ein Komma als Trenner
20 für Tausender verwendet (z.B. 100,000), und ein Punkt wird für Nachkommastellen verwendet (z.B. 1.6). Einige Autoren verwenden jedoch einen Punkt als Trenner für Tausender und ein Komma für Nachkommastellen. In einigen Sprachen, beispielsweise in Deutsch oder Niederländisch, wird ein Leerzeichen zwischen Tausendern und ein Komma für die Nachkommastellen ver-
25 wendet. Doch auch in diesen Sprachen wird häufig ein Punkt für Nachkommastellen verwendet. Um die korrekte Interpretation zu finden, müssen individuelle oder allgemeine Instruktionen durch den Benutzer gegeben werden. Alternativ wird der gesamte Text nach Zahlen durchsucht, und es wird anhand einer statistischen Analyse entschieden, welche Notation von dem Autor
30 benutzt worden ist. Normen für das Layout von Zahlen sind ebenfalls verfügbar.

Wenn das Wort "billion" als Angabe einer Zahl erkannt wird, muß der Benutzer bestimmen, ob es als 10^9 oder 10^{12} interpretiert werden soll.
35

Sehr große oder sehr kleine Zahlen werden häufig als Produkt aus einer Dezimalzahl und einer Zehnerpotenz geschrieben (z.B. $2,4 \times 10^5$). Solche Aus-

drücke werden ebenfalls als eine Zahl erkannt und interpretiert. Ebenso, wenn ein Toleranzbereich einer Zahl durch Ausdrücke wie " ± 5 " oder " (± 5) " angegeben ist, werden solche Ausdrücke als zu der vorangehenden Zahl gehörig betrachtet.

5

Nachdem eine vollständige Zahl festgestellt und identifiziert worden ist, wird überprüft, ob diese Zahl Teil einer Gruppe von zusammengehörenden Zahlen ist. Zu diesem Zweck werden die nachfolgenden Zeichenketten darauf überprüft, ob sie Zahlen angeben und durch Kommas, Semikolons, Bindestriche
10 oder Bindewörter wie "und", "oder", "bis" usw. getrennt sind. In den weiteren Verfahrensschritten werden die Zahlen einer solchen Gruppe als eine Gesamtheit behandelt.

Wenn eine Zahl (Gruppe von Zahlen) gefunden wurde, ist es wahrscheinlich,
15 daß die Zeichenkette unmittelbar hinter dieser Zahl eine Einheit bezeichnet. Es ist jedoch auch möglich, daß die Zahl eine dimensionslose Größe darstellt und nicht von einer Einheit begleitet ist. Aus diesem Grund wird in Schritt 14 versucht, die Zahl als eine solche ohne Dimension zu klassifizieren.

20 Zu diesem Zweck wird wenigstens einer der folgenden Tests ausgeführt, und die Zahl wird als dimensionslose Zahl ausgegrenzt, wenn wenigstens einer der Tests erfolgreich ist.

- Es wird überprüft, ob der Zahl das %-Zeichen folgt.

25

- Die Zeichenkette hinter und die Zeichenkette vor der Zahl werden mit einer Liste verglichen, die alle Namen von Monaten enthält, und die Zahl selbst wird darauf überprüft, ob sie ein Jahr repräsentiert. Es wird angenommen, daß eine Zahl ein Jahr repräsentiert, wenn auf vier Ziffern, von
30 denen die ersten beiden "19" sind, eine geschlossene Klammer oder ";" oder ":" folgt.

- Die Zeichenkette vor der Zahl wird mit einer Liste von Schlüsselwörtern wie "Tabelle", "Figur", "Kategorie" und dergleichen verglichen.

35

- Die Zeichenkette unmittelbar hinter der Zahl wird mit einem Wörterverzeichnis verglichen, um diese Zeichenkette als ein normales Wort zu

klassifizieren. Das Wörterverzeichnis ist so gestaltet, daß normale Wörter, die in der gleichen Weise wie eine Einheit geschrieben werden, ausgelassen oder speziell markiert sind.

- 5 Wenn sich ergibt, daß die Zahl eine dimensionslose Größe angibt, wird die Suche nach Zahlen mit Schritt 12 fortgesetzt. Anderfalls wird in Schritt 16 ein Korrekturflag zurückgesetzt.

2.2 Interpretation der Zeichenketten unmittelbar hinter der Zahl.

10

Wenn in Schritt 12 eine Zahl gefunden wurde und der Dimensionstest positiv ist, wird in Schritt 18 versucht, die Zeichenkette unmittelbar hinter dieser Zahl als eine Einheit zu interpretieren. Zu diesem Zweck wird die Zeichenkette mit einer Liste von Namen und Symbolen für Einheiten verglichen.

- 15 Wenn die Zeichenkette mit einem Namen oder einem Symbol einer Einheit identisch ist, wird sie als diese Einheit interpretiert, und die Dimension der Einheit wird in einem Register gespeichert. In einigen Fällen mag die Zeichenkette als solche mehrdeutig sein, so daß sie nicht zwingend als eine Einheit interpretiert werden muß. Berücksichtigt man jedoch, daß die Zeichen-
- 20 kette unmittelbar nach einer Zahl auftritt, ist die Interpretation als eine Einheit vernünftig.

- In ähnlicher Weise wird die Zeichenkette als eine Einheit interpretiert, wenn sie nur aus Symbolen für Einheiten besteht, die durch Malpunkte, Schrägstriche oder Potenzangaben getrennt sind, und die Dimension wird
- 25 aus den Dimensionen und Potenzen der einzelnen Faktoren bestimmt.

- Wenn die Zeichenkette erfolgreich als eine Einheit interpretiert werden konnte, wird die nächste Zeichenkette untersucht (Schritt 20) und der gleichen Interpretationsprozedur unterzogen. Auf diese Weise können Einheiten, die sich über mehr als eine Zeichenkette erstrecken (z.B. "N m" oder "miles per hour"), vollständig erfaßt werden. Die Dimension der vollständigen Einheit wird aus den Dimensionen abgeleitet, die den einzelnen Zeichenketten zugeordnet wurden.
- 30

35

Wenn eine Zeichenkette in Schritt 18 nicht als eine Einheit interpretiert werden konnte, wird auf diese Zeichenkette in Schritt 22 eine Korrekturpro-

zedur angewandt. Diese Korrekturprozedur entspricht im wesentlichen derjenigen, die in herkömmlichen Rechtschreibprozessoren angewandt wird. Es werden jedoch einige spezielle Vorkehrungen im Hinblick auf Fehler getroffen, die typischerweise in Verbindung mit der Schreibweise von Einheiten auftreten. Wenn eine vernünftige Korrektur verfügbar ist, wird in Schritt 24 das Korrekturflag gesetzt, und die Interpretationsprozedur nach Schritt 18 wird auf die korrigierte Zeichenkette angewandt.

Wenn andererseits eine Korrektur nicht möglich war, wird in Schritt 26 überprüft, ob die Zeichenkette, die nicht interpretiert werden konnte, die erste Zeichenkette hinter der Zahl war. Wenn sie nicht die erste Zeichenkette war, wird gefolgert, daß eine vollständige Angabe einer Einheit gefunden worden ist. Wenn sie die erste Zeichenkette war, wird in Schritt 28 das Korrekturflag gesetzt, und die Suche nach Zahlen wird fortgesetzt (Schritt 12).

15

3. Prüfung auf Zulässigkeit der Einheit

Wenn eine Angabe einer Einheit gefunden wurde, wird in Schritt 30 überprüft, ob die Dimension dieser Einheit, die bei der Interpretationsprozedur bestimmt worden ist, in einer Liste von Dimensionen, die in dem Anwendungsgebiet tatsächlich auftreten, gefunden werden kann. Zusätzlich wird die Umgebung der Einheit (z.B. der Satz, in dem die Einheit auftritt, oder eine vorgegebene Anzahl von Zeichenketten vor und hinter der Einheit) nach Schlüsselwörtern oder Symbolen durchsucht, die sich auf Größen beziehen. Wenn ein solches Schlüsselwort oder Symbol gefunden wird, wird die Dimension der Größe mit der Dimension der festgestellten Einheit verglichen. Wenn die Dimension der Einheit in dem Anwendungsgebiet nicht auftritt oder nicht mit der Dimension der zugehörigen Größe konsistent ist, wird erneut versucht, die zu der Einheit gehörende Zeichenkette zu korrigieren (Schritt 22).

30

Wenn die Dimensionsüberprüfung in Schritt 30 erfolgreich war, wird in Schritt 32 geprüft, ob die Einheit und ihre Bezeichnung nach der SI-Norm zulässig sind. Bei dieser Prüfung werden die folgenden Kriterien überprüft:

35

- Sind alle Symbole, die in der Einheit auftreten, nach der SI-Norm zulässig?

- Ist die Schreibweise als ein Produkt zulässig oder müssen einige oder alle Faktoren durch spezielle Symbole ersetzt werden (z.B. "W" anstelle von "J·s⁻¹")?

5 - Entspricht die Verwendung von Präfixen der SI-Norm?

Wenn wenigstens eine dieser Anforderungen nicht erfüllt ist, wird eine entsprechende zulässige Einheit oder Bezeichnung in Schritt 34 ausgewählt. Dann wird in Schritt 36 überprüft, ob die Änderung der Einheit oder Bezeichnung einen Einfluß auf den Zahlenwert der Größe hat. Erforderlich je-

10 denfalls wird ein neuer Zahlenwert für jede der Zahlen berechnet, die zu der in Schritt 12 festgestellten Gruppe gehören.

- Die Berechnung der neuen Zahlenwerte wird in Schritt 38 ausgeführt. Ein
- 15 Satz von Umwandlungsformeln, die die Beziehung zwischen den verbotenen Einheiten und den zulässigen SI-Einheiten bestimmen, sind in einer Tabelle gespeichert. Im allgemeinen besteht die Umwandlung einfach in einer Multiplikation mit einem bestimmten Faktor.

- 20 Es ist nicht notwendig, Umwandlungsfaktoren für jedes mögliche Paar von Einheiten zu speichern. Wenn die festgestellte Einheit eine Potenz einer verbotenen Einheit oder ein Produkt von Potenzen von verbotenen Einheiten ist, kann sie allgemein durch eine entsprechende Potenz oder ein Produkt von Potenzen von zulässigen Einheiten ersetzt werden. In diesem Fall werden
- 25 die Umwandlungsfaktoren zu derselben Potenz genommen wie die zugehörigen Einheiten, und der Zahlenwert wird mit dem Produkt der Umwandlungsfaktoren multipliziert. Zum Beispiel wird die verbotene Einheit "pound-force per square inch" (lbf in⁻²) in die SI-Einheit "Pascal" (Pa = N m⁻²) umgewandelt. Der Umwandlungsfaktor für "lbf" in "N" ist 4,448, und der Um-
- 30 wandlungsfaktor für "in" in "m" ist 0,0254. Folglich ist der Gesamt-Umwandlungsfaktor gegeben durch:

$$4,448 \cdot (0,0254)^{-2} = 6,8944, \text{ so daß die Größe}$$

- 35 "1 lbf in⁻²" umgewandelt wird in "6,8944 Pa".

In der wissenschaftlichen Literatur gibt die Anzahl von Dezimalstellen hinter

dem Dezimalkomma im allgemeinen die Genauigkeit an, mit der die entsprechende Größe bestimmt ist. Zum Beispiel bedeutet die Bezeichnung "0,3 kg", daß die Toleranz in der Größenordnung von 10 g liegt, während die Bezeichnung "0,300 kg" angibt, daß die Toleranz nur in der Größenordnung von einigen Zehntel Gramm liegt. Somit ist strenggenommen eine Umwandlung wie

$$0,1 \text{ lb} = 0,04536 \text{ kg}$$

irreführend, obgleich der Zahlenwert korrekt ist. In diesem Beispiel gibt "0,1 lb" eine Toleranz von $\pm 0,05 \text{ lb}$ an, was etwa $\pm 0,03 \text{ kg}$ entspricht. Folglich sollte der Zahlenwert für kg auf 0,045 gerundet werden. Aus diesem Grund werden die in Schritt 38 berechneten Zahlenwerte so gerundet, daß die durch die letzte Dezimalstelle hinter dem Dezimalkomma angegebene Toleranz der durch die ursprüngliche Zahl angegebenen Toleranz entspricht.

15

Wenn die in Schritt 38 berechneten neuen Zahlenwerte viele Null-Digitalstellen vor oder hinter dem Dezimalkomma aufweisen, werden Präfixe (für "milli", "kilo", "mega" und dergleichen in Übereinstimmung mit der SI-Norm zu der Einheit hinzugefügt oder davon entfernt, und der Zahlenwert wird mit einer entsprechenden Zehnerpotenz multipliziert oder dividiert.

20

Alternativ wird vorgeschrieben, daß beispielsweise die Einheit "Meilen" stets durch "Kilometer" ersetzt wird, während die Einheit "Inch" (Zoll) stets durch "Zentimeter" ersetzt wird, so daß große Umwandlungsfaktoren vermieden werden. Dieses Verfahren ist besonders vorteilhaft, wenn indirekte Zahlenangaben in Schritt 12 erkannt wurden. In diesem Fall wird beispielsweise "einige Zoll" umgewandelt in "einige Zentimeter" und nicht in "einige Meter".

25

30 4. Korrektur des Layouts

Wenn in Schritt 32 gefunden wurde, daß die Einheit und ihre Bezeichnung zulässig sind, oder wenn in Schritt 36 entschieden wurde, daß der Zahlenwert (die Zahlenwerte) nicht geändert zu werden braucht, oder wenn die neuen Zahlenwerte in Schritt 38 berechnet worden sind, werden die SI-Einheit und die zugehörigen Zahlen in der zugelassenen Schreibweise und in Übereinstimmung mit dem vorgewählten Layout (Schritt 40) geschrieben

35

und mit den ursprünglichen Zeichenketten verglichen (Schritt 42). Wenn die neuen Zeichenketten, die in Schritt 40 geschrieben wurden, nicht mit den ursprünglichen Zeichenketten identisch sind, wird in Schritt 44 ein Korrekturflag gesetzt.

5

5. Korrektur

10 In Schritt 46 wird festgestellt, ob das Korrekturflag zu irgendeiner Zeit während der oben beschriebenen Prozedur gesetzt worden ist. Wenn das Korrekturflag nicht gesetzt worden ist, wird die Suche nach Zahlen unmittelbar mit der nächsten Zeichenkette des Textes fortgesetzt (Schritt 12). In diesem Fall ist die "nächste" Zeichenkette diejenige Zeichenkette, die in Schritt 18 nicht als eine Einheit interpretiert werden konnte.

15 Wenn eine vollständige Zeile des Textes durch Wiederholung der Schritte 12 bis 46 durchsucht worden ist und wenn keine Korrektur in dieser Zeile erforderlich war, so wird die Zeile beispielsweise auf einem Bildschirm angezeigt.

20 Wenn in Schritt 46 ein Korrekturflag festgestellt wird, wird in Schritt 48 eine Korrekturprozedur ausgeführt. In Abhängigkeit von einer Instruktion des Benutzers kann die Korrektur automatisch oder durch Interaktion mit dem Benutzer ausgeführt werden.

25 Im Fall der interaktiven Korrektur wird die Zeile, in der der Fehler festgestellt wurde, dem Benutzer angezeigt, und die Zeichenketten, die korrigiert werden müssen, werden hervorgehoben. Zusätzlich wird die vorgeschlagene Korrektur, d.h., die in Schritt 40 geschriebene Folge von Zeichenketten, angezeigt. Dann wird durch den Benutzer entschieden, ob der ursprüngliche
30 Text ohne Korrektur beibehalten oder die vorgeschlagene Korrektur ausgeführt werden soll, oder der Benutzer kann eine andere Korrektur eingeben.

Der korrigierte Text wird in einer Ausgabedatei gesichert.

35 Im Fall der automatischen Korrektur wird die Korrektur gemäß Schritt 40 automatisch in der Ausgabedatei gesichert, und die alten Zeichenketten, die gelöscht worden sind und die neuen Zeichenketten, die dafür eingesetzt

worden sind, werden durch geeignete Zeichen markiert. Wenn keine Korrektur verfügbar ist, beispielsweise, wenn die Dimensionsüberprüfung in Schritt 30 ergibt, daß die Einheit in dem Anwendungsgebiet nicht zu finden ist, selbst wenn alle Korrekturmöglichkeiten überprüft worden sind, und
5 wenn die korrekte Dimension nicht aus dem Kontext bestimmt werden kann, wird keine Korrektur ausgeführt, und ein Fehlersignal wird dem Benutzer angezeigt.

In einer modifizierten Ausführungsform des Korrekturprozesses hat der Benutzer die Möglichkeit, bestimmte Arten von Fehlern zu bestimmen, die
10 automatisch korrigiert werden sollen, während andere Fehler interaktiv korrigiert werden. In einer noch weiter verfeinerten Version kann ein Lernprozeß installiert sein, so daß, wenn der Benutzer einmal einer bestimmten Korrektur zugestimmt oder diese veranlaßt hat, Fehler derselben Art automa-
15 tisch analog zu dieser Korrektur gehandhabt werden.

20

25

30

35

P 37 50 135.6-08

PATENTANSPRÜCHE

- 5 1. Verfahren zur Prüfung des richtigen und konsistenten Gebrauchs von fehlerträchtigen Datensegmenten in einer Kette von Datensegmenten in einem Textverarbeitungssystem, mit den Schritten:
- (a) Speichern einer Liste von Symbolen, die fehlerträchtige Datensegmente repräsentieren,
- 10 (b) Speichern von Kriterien, wie etwa typischen syntaktischen Strukturen in der Bezeichnung oder in dem Kontext, in dem fehlerträchtige Datensegmente auftreten, die festlegen, ob ein Symbol oder eine Kombination von Symbolen als ein fehlerträchtiges Datensegment interpretiert wird oder nicht,
- (c) Spezifizieren von zulässigen fehlerträchtigen Datensegmenten und zuläs-
- 15 sigen Bezeichnungen derselben,
- (d) Durchsuchen des Textes nach Zeichenketten, die irgendwelche der aufgelisteten Symbole enthalten, die als ein fehlerträchtiges Datensegment oder Teil eines fehlerträchtigen Datensegments interpretiert werden können, anhand der gespeicherten Kriterien,
- 20 (e) wenn eine solche Zeichenkette gefunden wird, vergleichen des zugehörigen fehlerträchtigen Datensegments und der vollständigen Bezeichnung desselben mit den in Schritt c) spezifizierten fehlerträchtigen Datensegmenten und Bezeichnungen und
- (f) Erzeugen eines Fehlersignals, wenn das fehlerträchtige Datensegment
- 25 oder die Bezeichnung nicht zulässig ist, dadurch **gekennzeichnet**, daß die fehlerträchtigen Datensegmente Einheiten sind und die in Schritt b) gespeicherten Kriterien eine Überprüfung einschließen, ob in einer vorgegebenen Umgebung der Zeichenkette entweder eine Zahlenangabe oder ein Schlüsselwort oder Symbol gefunden wird, das sich auf eine Größe bezieht.
- 30 2. Verfahren nach Anspruch 1, bei dem zulässige Einheiten spezifiziert werden, indem eine Liste von Symbolen für zulässige Einheiten sowie Bildungsregeln für aus solchen Symbolen gebildete mögliche mathematische Ausdrücke (z.B. Produkte) gespeichert werden.

35

3. Verfahren nach Anspruch 2, bei dem für jede in Schritt d) gefundene Einheit eine Dimension berechnet wird und eine Einheit als unzulässig betrachtet wird, wenn ihre Dimension nicht in dem Anwendungsgebiet auftritt oder von der Dimension der Größe verschieden ist, die durch ein Symbol
5 oder Schlüsselwort angegeben wird, auf das sich die Einheit in dem Text bezieht.
4. Verfahren nach einem der Ansprüche 1 bis 3, bei dem in Schritt d) der Text zunächst nach Zahlen durchsucht wird und die Suche nach Einheiten
10 auf die Zeichenketten unmittelbar hinter einer Zahl oder einer zuvor gefundenen Einheit beschränkt wird.
5. Verfahren nach einem der Ansprüche 1 bis 4, bei dem in Schritt d) die Zeichenketten mit einem Wörterverzeichnis verglichen werden, aus dem alle
15 Wörter entfernt worden sind, die dieselbe Schreibweise wie eine Einheit haben, und bei dem eine Zeichenkette ohne Überprüfung weiterer Kriterien zurückgewiesen wird, wenn sie mit einem Wort des Wörterverzeichnisses identifiziert werden kann.
- 20 6. Verfahren nach einem der vorstehenden Ansprüche, bei dem eine Zeichenkette, die in Schritt d) nicht als eine Einheit interpretiert werden konnte, einer Rechtschreibüberprüfung und einem Korrekturprozeß unterzogen wird und erneut anhand der Interpretationskriterien überprüft wird, wenn die Rechtschreibkorrektur ausgeführt worden ist.
- 25 7. Verfahren nach einem der vorstehenden Ansprüche, bei dem Schritt c) einen Schritt der Speicherung einer entsprechenden zulässigen Einheit für jede unzulässige Einheit enthält und bei dem, wenn eine Einheit oder Bezeichnung in Schritt e) als unzulässig gefunden wurde, die entsprechende zu-
30 lässige Einheit und/oder Bezeichnung bestimmt wird.
8. Verfahren nach Anspruch 7, mit den Schritten
 - Speichern von Umwandlungsformeln zur Umwandlung unzulässiger Ein-
35 heiten in zulässige,
 - Identifizieren der Zahlenwerte aller Zahlen, ob sie einer Einheit zugeord-

net sind oder nicht und

- Berechnen neuer Zahlenwerte gemäß der Umwandlungsformel für die unzulässige Einheit und die entsprechende zulässige Einheit.

5

9. Verfahren nach Anspruch 8, bei dem festgestellt wird, daß eine Zahl der Einheit zugeordnet ist, wenn sie der Bezeichnung für die Einheit unmittelbar vorangeht oder wenn sie mit einer anderen Zahl, die der Einheit zugeordnet ist, nur durch Interpunktionszeichen wie Komma, Semikolon, Bindestriche
10 oder durch Wörter wie "und", "oder", "bis" verbunden ist.

10. Verfahren nach Anspruch 8, bei dem eine Zahl zu derselben syntaktischen Gruppe gehört oder Bestandteil eines Satzes ist.

- 15 11. Verfahren nach Anspruch 10, 9 oder 8, mit den zusätzlichen Schritten:

- Bestimmen der Toleranz infolge des Rundungsfehlers der Zahl,
- Berechnen der entsprechenden Toleranz des neuen Zahlenwertes gemäß
20 der Umwandlungsformel und
- Runden des neuen Wertes auf die größte Dezimalstelle, für die der Rundungsfehler kleiner ist als die berechnete Toleranz.

- 25 12. Textverarbeitungssystem zur Prüfung des richtigen und konsistenten Gebrauchs von Einheiten in einem gegebenen Text, nach einem der vorstehenden Ansprüche.

30

35

